BBN TECHNOLOGIES
A Verizon Company

EARS The Blue Team
BBN LIMSI UPitt UW

BBN+ Rich Transcription System for CTS

Amit Srivastava, Daben Liu, Francis Kubala,
Daniel Kiecza, Jared Maguire, Rich Schwartz
BBN Technologies

Matthew Snover, Bonnie Dorr
University of Maryland, College Park

Mari Ostendorf, Jay Kim, Sarah Schwarm, Bill McNeill
University of Washington

RT-03F Workshop

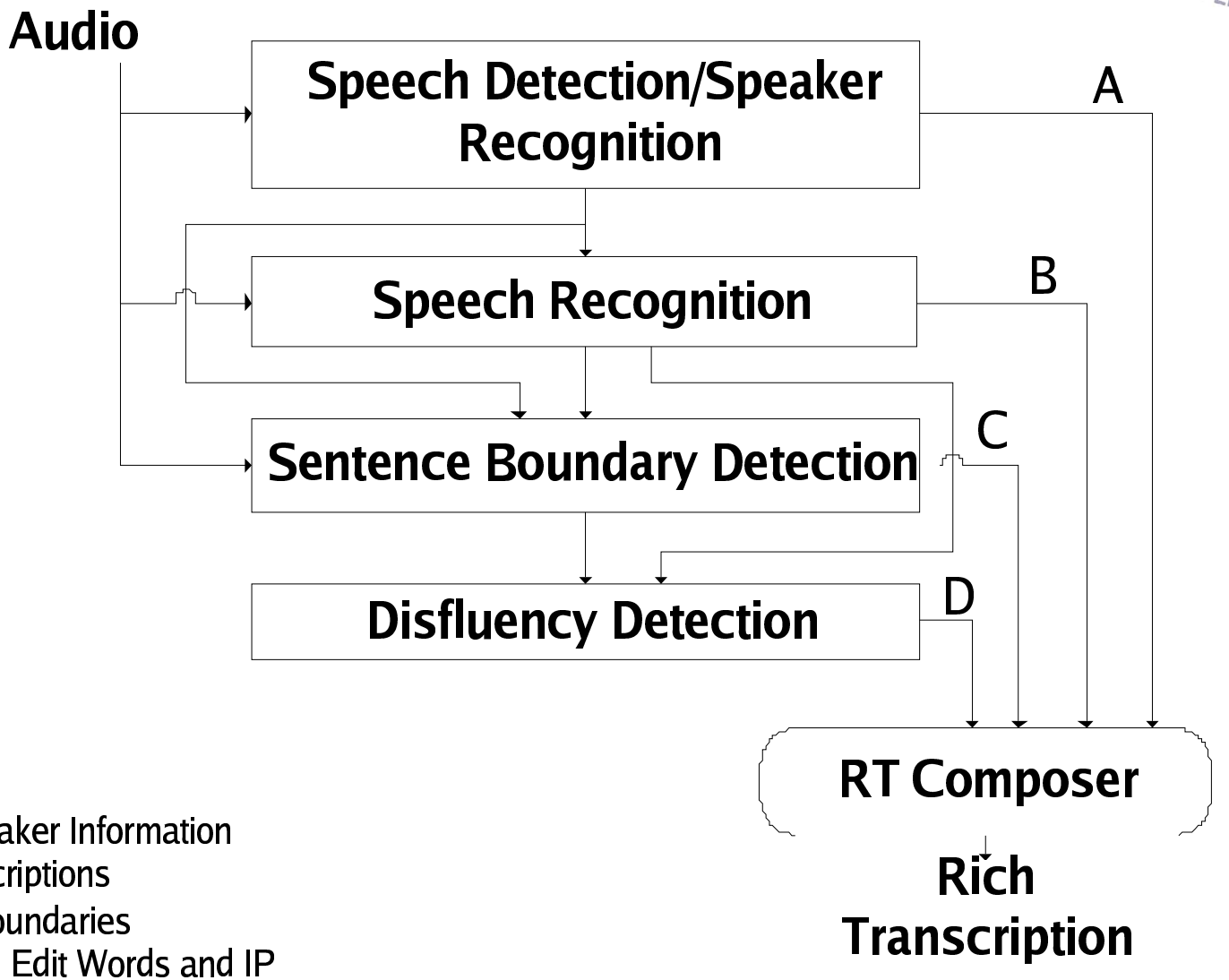
Washington, DC

13 November, 2003

Outline

- **Rich Transcription System for CTS**
  - Overview
  - Speech Detection and Speech-to-Text
  - BBN Sentence Boundary Detection
  - SBD System Combination
  - Progress Results
- **Evaluation Results**
- **Post-Eval Progress on Sentence Boundary Detection**
- **Summary**

CTS Rich Transcription System Overview



A = Speech/Speaker Information
B = Word Transcriptions
C = Sentence Boundaries
D = Filler Words, Edit Words and IP

Speech Detection & Speech-to-Text

- **Acoustic Segmentation using Cross-Channel Event Modeling[1]**
  - **2 classes, *Non-Speech (N)*, and *Speech (S)*, for each channel**
  - **GMMs trained for each of the 4 Cross-Channel Events: *NN, NS, SN, SS***
  - **Viterbi Decoding to label each Data Frame with an Event**
  - **Silence or Non-Speech Segments are discarded**

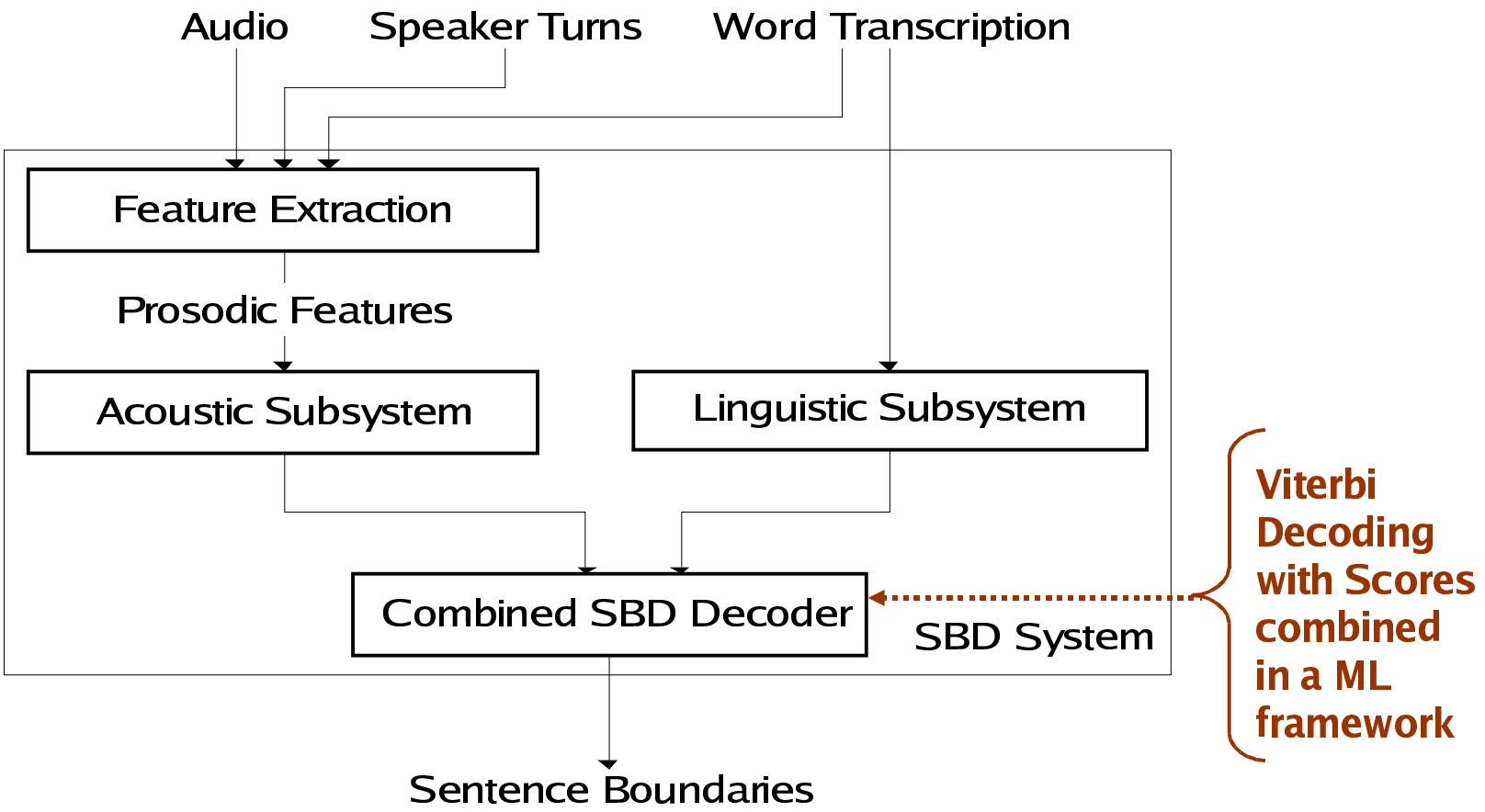
- **Single best BBN-only system used in RT03 Spring evaluation**
  - **Time constraint and prioritization**
  - **No pause-fillers in BBN + LIMSI combined STT output**

- **BBN + LIMSI Combined STT Primary Submission for RT03S was used for Metadata experiments post-RT03F Evaluation**

[1] Daben Liu, Francis Kubala, "A Cross-Channel Modeling Approach for Automatic Segmentation of Conversational Telephone Speech," ASRU'2003, to be presented, US Virgin Island, Dec 2003.

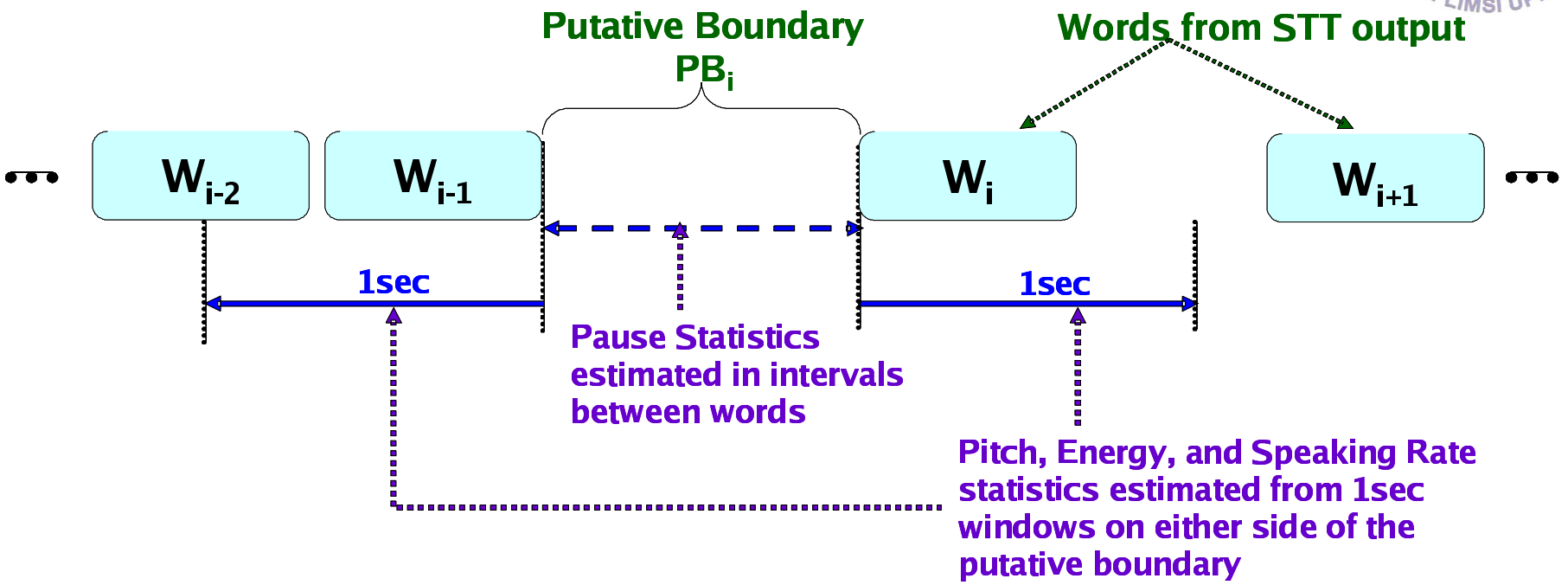
[2] R. Schwartz, et al., "Speech recognition in multiple languages and domains: The BBN/LIMSI EARS system," Proc. ICASSP-2004, Montreal, Canada, May 2004, appeared elsewhere on this proceeding.

BBN Sentence Boundary Detection System



- **Sentence boundaries hypothesized at each word boundary**

Prosodic Feature Extraction



Feature Type	Feature Description	# Features
Pause	• Pause Duration • Pause Attribute (Filler, Breath, etc.) • Time since last Pause • Normalized Pause Duration	10
Speaking Rate	• Absolute Value • Difference across Putative Boundary	2
Energy	• Absolute Values • Difference across Putative Boundary • First Difference of Energy	6
Pitch	• Discontinuous Chains in Voiced Regions • Interpolated Continuous Pitch • First-Order Pitch Differences	30
	Total	48

Acoustic Subsystem for SBD

- **Acoustic Subsystem**
  - 2-layer feed-forward neural network trained on 48 prosodic features
  - 3 Sentence classes: *statement*, *incomplete*, *no-sentence*
  - Features are discrete, continuous, and boolean
  - #Nodes: 48 input, 500 hidden, 3 output
  - BackProp training, minimum cross-entropy error criterion
  - 50 hours of LDC corpus used for training
  - NN scores are estimates of posterior probability of sentence class
    - Class likelihoods estimated by scaling with the class priors

Linguistic Subsystems for SBD

- **Word-based Linguistic Subsystem**
  - LM probabilities estimated by the BBN BYBLOS LM tools
  - Trigram LM with Sentence-class tokens inserted between words
  - LM trained on 500K words from the LDC 50 hour training trans.
  - LM scores used as transition probabilities in Combined Viterbi Decoder

- **Hybrid Word-POS Linguistic Subsystem**
  - Starts from same resources and tools as the Word-based system
  - Ratnaparkhi[3] MAXENT tagger used to hypothesize POS tags
  - Top 1000 frequent words left as is in the transcriptions
  - Rest of the words are replaced by their POS tags
  - Trigram LM estimated with Hybrid word-POS transformed transcripts

[3] A. Ratnaparkhi, "A Maximum Entropy Part-of-Speech Tagger," in Proc. of the Empirical Methods in Natural Language Processing Conference, 1996, pp. 133-141.

SBD Progress Summary for Dev03F

- **Word-based SBD System**

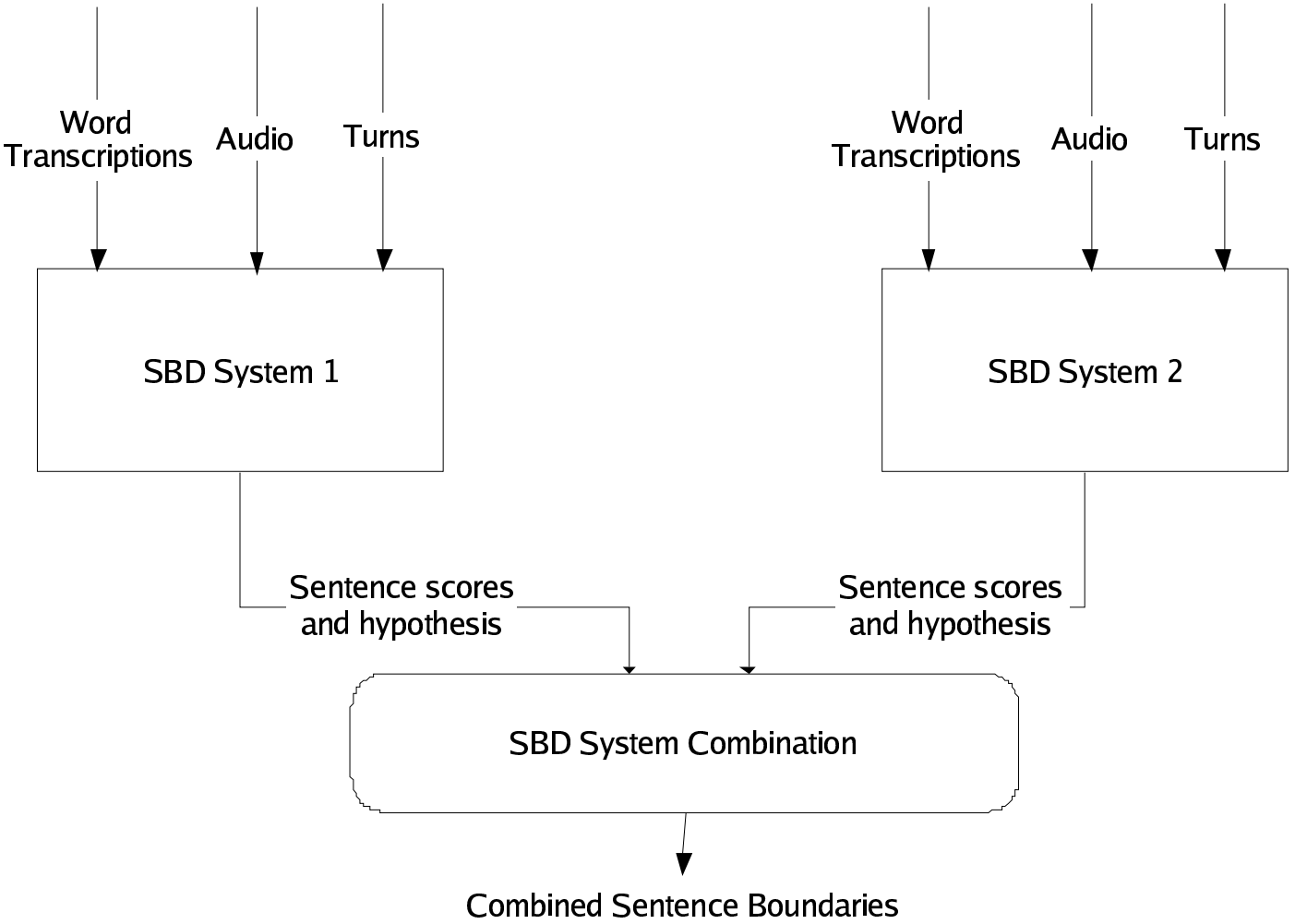
A.SBD Experiment	SER
August DryRun baseline (old training)	63.0
With compound words	60.8
Without compound words	52.5
100 frames feature window size	52.0
System tuning using class biasing	51.6

- **Hybrid Word-POS based SBD System***

SBD Experiment	SER
Pure POS LM	69.7
Hybrid Word-POS LM	52.3

* **Bug in the POS Training Procedure discovered post-evaluation**

SBD System Combination



SBD System Combination Overview

- **SBD System Combination with UW**
  - BBN Word-based SBD output combined with UW SBD output
    - Sentence class scores and SBD hypothesis at each word boundary used to create a 9 dimensional score vector
    - Sentence boundaries from reference transcriptions are transferred to word boundaries in STT hypothesis
    - NN with 50 hidden nodes trained using MSE backprop
  - During development, training and test are Jacknifed on 2 equal halves of the Dev03F set
    - For the Evaluation, SBD Combination NN is trained on the complete Dev03F set
- **System Combination of BBN SBD Systems**
  - Same Acoustic subsystem used with Word-based and Hybrid Word-POS based Linguistic Subsystems
  - Same NN-based procedure used for SBD system combination

SBD System Combination Results on Dev and Eval

- **BBN1 + UW**

SBD Experiment	Dev03F SER	Eval03F SER
UW Serial System	50.6	46.6
BBN Word-based SBD System	51.6	50.9
SBD Combination	49.0	46.7

- **BBN1 + BBN2**

SBD Experiment	Dev03F SER	Eval03F SER
Word-based system	51.6	50.9^{+}
Hybrid Word-POS based system	52.3	48.7*
SBD Combination	51.1	49.3^{x}

+ **Primary Submission**

* **Did not submit this system in the Evaluation due to non-positive results on Dev03F**

x **Contrast Submission**

Final Results

System	SBD	Edit	Filler	IP	SR	RT1	RT03
Dev03F							
BBN + UMD	51.6	85.4	49.5	70.5	11.4	27.2	37.3
Eval03F							
UW + BBN	46.7	88.5	51.0	68.4	10.1	25.2	33.9
BBN + UMD	50.9	87.9	48.8	69.0	10.2	25.3	34.7

- **Results were consistent going from Dev to Eval**
  - **Eval is an easier test set compared to Dev**
  - **Most of the gain in RT03 TER is due to better RT1 TER**

Post-Eval Experiments

- **Processed the BBN + LIMSI primary submission STT output for MDE experiments**
  - **Re-inserted pause-fillers from single-best BBN STT output into the Combined STT output**
- **Bug fixes in the Hybrid Word-POS based SBD system**

BBN SBD Improvements Post-Eval on Dev03F

- **Word-based SBD System**

SBD Experiment	RT1 TER	SBD SER
Pre-Eval System on Single-best STT	27.2	51.6
Pre-Eval System on BBN+LIMSI STT	24.3	47.2

- **Hybrid Word-POS based SBD System**

SBD Experiment	RT1 TER	SBD SER
Pre-Eval System on Single-best STT	27.2	52.3
Pre-Eval System on BBN+LIMSI STT	24.3	47.0
With POS bug fixes on new STT	24.3	46.3

Summary

- BBN combined with UW for Sentences
  - UW used the combined sentences to re-run TBL for Disfluencies
- BBN integrated with UMD for Disfluencies
  - Disfluency hypothesis from UMD integrated into a single RTXML file
- Despite insufficient time, BBN developed 2 systems for Sentence Boundary Detection in CTS
  - Hybrid Word-POS based SBD system (shows a lot of promise)
  - System combination (needs more work)
- Future Work
  - Combine output from BBN, UW, and UMD into an integrated RT output
  - Better SBD system combination techniques
  - Integrate Parsing for Rich Transcription